REMARKS

The Office Action has been carefully considered and the Examiner's comments duly noted. The Reconsideration of this application in the light of the amendments for the specification, amendments for the claims is respectfully solicited.

A separate request for three-months term extension has been submitted. If for some reason this become the attached, please use this as your authorization to charge the account 50-1529.

It is noted that Claims 1 to 14, 19 to 22 and 24 to 28 would be allowable. The claims now in this application have been amended to overcome the Examiner's formal objections.

There are terms which the Examiner has objected to in the Office Action, and for this purpose applicants respectfully submit the enclosed papers which indicate the terms as being International standardization in the MPEG group. These sheets are headed International Organization For Standardization, Organisation Internationale DE Normalisation, and more specifically referred to as ISO/IEC JTC 1/SC 29/WG 11 N28O2 with the date of Vancouver, July 1999.

If there are any points outstanding the Examiner is respectfully asked to call applicant's attorney in order to do what is necessary to place the application into condition for allowance.

In view of the foregoing, early and favorable reconsideration in this application is respectfully solicited.

Respectfully submitted,

J. Harold Nissen
Reg. No. 17,283

Enclosures:
Appendix - claims with brackets and underlining
International Organization For Standardization
ISO/IEC JTC 1/SC 29/WG 11 N28O2

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11 N2802
Vancouver, July 1999

Information technology - Generic coding of audio-visual objects - Part 2: Visual

ISO/IEC 14496-2 FPDAM 1
Final Proposed Draft Amendment 1

Table V2 - 39 -- Probability for arithmetic decoding of spatial scalable binary shape

threshold = 2^{15}

2^{16}

```
USInt SI_bab_type_prob[1]={ 59808 }

USInt enh_intra_v_prob[128]={
 65476, 64428, 62211, 63560, 52253, 58271, 38098, 31981, 50087, 41042,
 54620, 31532,  8382, 10754,  3844,  6917, 63834, 50444, 50140, 63043,
 58093, 45146, 36768, 13351, 17594, 28777, 39830, 38719,  9768, 21447,
 12340,  9786, 60461, 41489, 27433, 53893, 47246, 11415, 13754, 24965,
 51620, 28011, 11973, 29709, 13878, 22794, 24385,  1558, 57065, 41916,
 25259, 55117, 48064, 12960, 19929,  5937, 25730, 22366,  5204, 32865,
  3415, 14814,  6634,  1155, 64444, 62907, 56337, 63144, 38112, 56527,
 40247, 37088, 60326, 45675, 51248, 15151, 18868, 43723, 14757, 11721,
 62436, 50971, 51738, 59767, 49927, 50675, 38182, 24724, 48447, 47316,
 56628, 36336, 12264, 25893, 24243,  5358, 58717, 56646, 48302, 60515,
 36497, 26959, 43579, 40280, 54092, 20741, 10891,  7504,  8109, 30840,
  6772,  4090, 59810, 61410, 53216, 64127, 32344, 12462, 23132, 19270,
 32232, 24774,  9615, 17750,  1714,  6539,  3237,  152
}

USInt enh_intra_h_prob[128]={
65510, 63321, 63851, 62223, 64959, 62202, 63637, 48019, 57072, 33553,
37041,  9527, 53190, 50479, 54232, 12855, 62779, 63960, 49604, 31847,
57591, 64385, 40657,  8402, 33878, 54743, 17873,  8707, 34470, 54322,
16702,  2192, 58325, 48447,  7345, 31317, 45687, 44236, 16685, 24144,
34327, 18724, 10591, 24965,  9247,  7281,  3144,  5921, 59349, 33539,
11447,  5543, 58082, 48995, 35630, 10653,  7123, 15893, 23830,   800,
3491, 15792,  8930,   905, 65209, 63939, 52634, 62194, 64937, 53948,
60081, 46851, 56157, 50930, 35498, 24655, 56331, 59318, 32209,  6872,
59172, 64273, 46724, 41200, 53619, 59022, 37941, 20529, 55026, 52858,
26402, 45073, 57740, 55485, 20533,  6286, 64286, 55438, 16454, 55656,
61175, 45874, 28536, 53762, 58056, 21895,  5482, 39352, 32635, 21633,
2137,  4016, 58490, 14100, 18724, 10461, 53459, 15490, 57992, 15126,
12034,  4340,  6761,  1859,  5794,  6785,  2412,  35
}
```

INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11 N2802
Vancouver, July 1999

Information technology – Generic coding of audio-visual objects – Part 2: Visual

ISO/IEC 14496-2 FPDAM 1

Final Proposed Draft Amendment 1

1.1.1.1.1 Border formation

When decoding a BAB, pixels from neighbouring BABs shall be used to make up the context. For both the INTRA and INTER cases, a 2 pixel wide border about the current BAB is used where pixels values are known, as depicted in Figure 7-11.

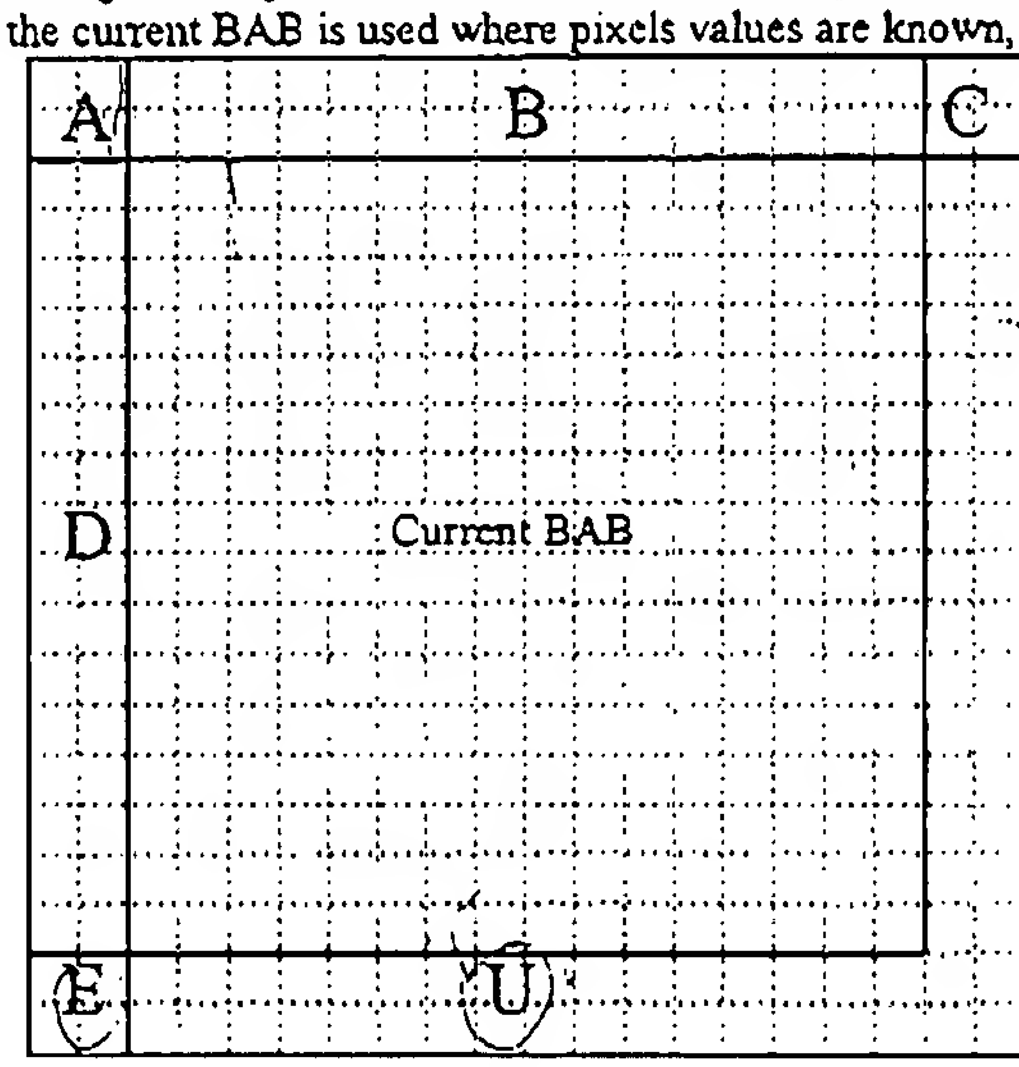


Figure 7-11 -- Bordered BAB. A: TOP_LEFT_BORDER. B: TOP_BORDER. C: TOP_RIGHT_BORDER. D: LEFT_BORDER. E: BOTTOM_LEFT_BORDER. U: pixels which are unknown when decoding the current BAB

If the value of conv_ratio is not equal to 1, a sub-sampling procedure is further applied to the BAB borders for both the current BAB and the motion compensated BAB.
The border of the current BAB is partitioned into 5 regions:

- TOP_LEFT_BORDER, which contains pixels from the BAB located to the upper-left of the current BAB and which consists of 2 lines of 2 pixels
- TOP_BORDER, which contains pixels from the BAB located above the current BAB and which consists of 2 lines of 16 pixels
- TOP_RIGHT_BORDER, which contains pixels from the BAB located to the upper-right of the current BAB and which consists of 2 lines of 2 pixels
- LEFT_BORDER, which contains pixels from the BAB located to the left of the current BAB and which consists of 2 columns of 16 pixels
- BOTTOM_LEFT_BORDER, which contains pixels from the BAB located to the bottom-left of the current BAB and which consists of 2 lines of 2 pixels

The TOP_LEFT_BORDER and [illegible]_RIGHT_BORDER are not sub-sampled, and k[illegible]s they are. The TOP_BORDER and LEFT_BORDER are sub-sample[illegible]ch as to obtain 2 lines of 16/conv_ratio pixels an[illegible] columns of 16/conv_ratio pixels, respectively.
The sub-sampling procedure is performed on a line-basis for TOP_BORDER, and a column-basis for LEFT_BORDER. For each line (respectively column), the following algorithm is applied: the line (respectively column) is split into groups of conv_ratio pixels. For each group of pixels, one pixel is associated in the sub-sampled border. The value of the pixel in the sub-sampled border is OPAQUE if half or more pixels are OPAQUE in the corresponding group. Otherwise the pixel is TRANSPARENT.
The 2x2 BOTTOM_LEFT_BORDER is filled by replicating downwards the 2 bottom border samples of the LEFT_BORDER after the down-sampling (if any).

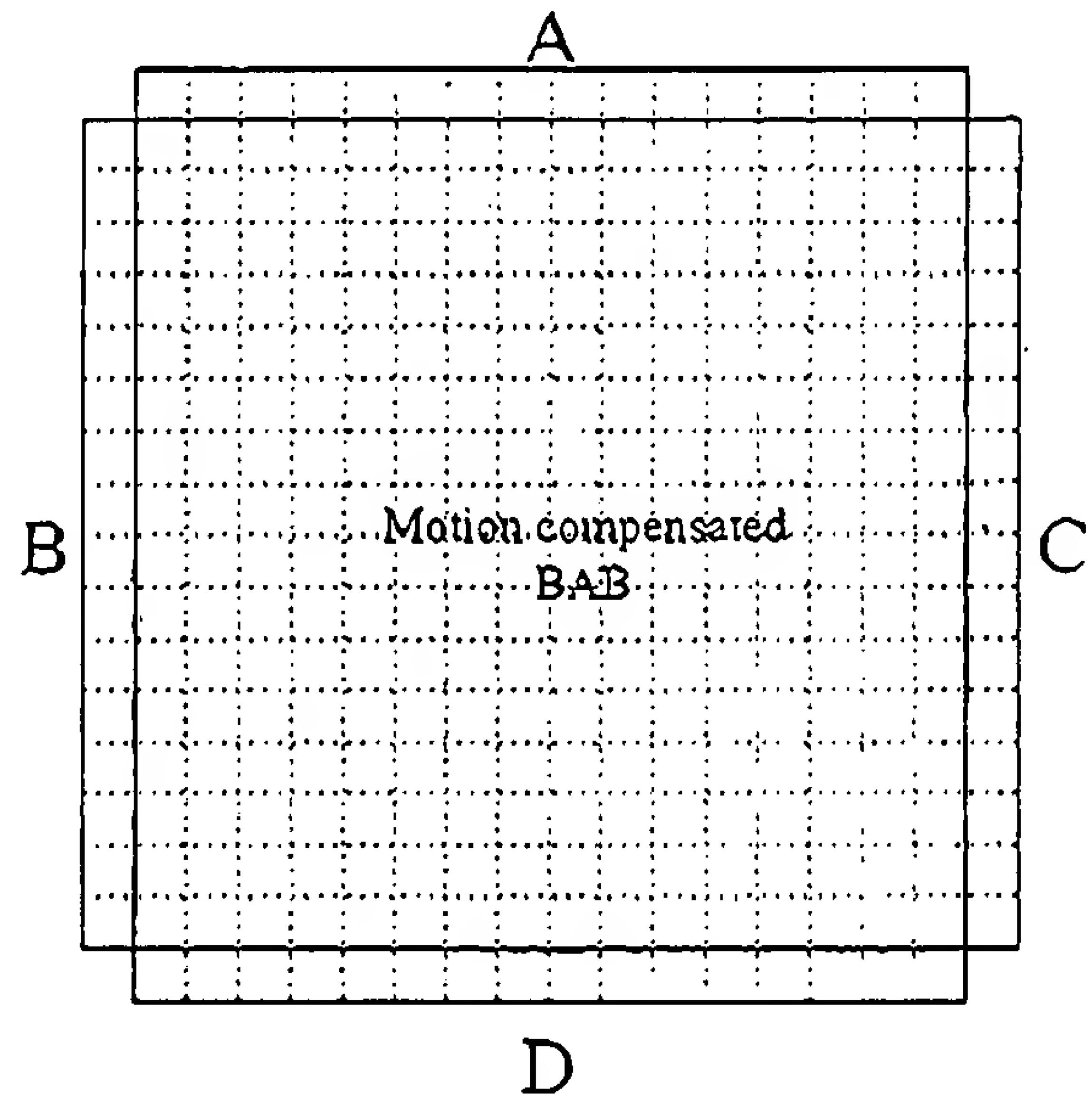


Figure 7-12 -- Bordered motion compensated BAB. A: TOP_BORDER. B: LEFT_BORDER. C: RIGHT_BORDER. D: BOTTOM_BORDER

In the case of a motion compensated BAB, the border is also partitioned into 4, as shown Figure 7-12:

- TOP_BORDER, which consists of a line of 16 pixels
- LEFT_BORDER, which consists of a column of 16 pixels
- RIGHT_BORDER, which consists of a column of 16 pixels
- BOTTOM_BORDER, which consists of a line of 16 pixels

ERNATIONAL ORGANISATION FOR STA RDISATION
GANISATION INTERNATIONALE DE NO LISATION
ISO/IEC JTC 1/SC 29/WG 11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC 1/SC 29/WG 11 N2802
Vancouver, July 1999

Information technology – Generic coding of audio-visual objects – Part 2: Visual

---

ISO/IEC 14496-2 FPDAM 1

Final Proposed Draft Amendment 1

1.1.1 Spatial scalable binary shape decoding

1.1.1.1 Decoding of base layer

The decoding process of the base layer is the same as non-scalable binary shape decoding process.

1.1.1.2 Decoding of enhancement layer

When spatial scalability is enabled (scalability is set to 1 and hierarchy_type is set to 0) with enhancement_type == 0 or When spatial scalability is enabled with enhancement_type == 1 and use_ref_shape == 0, scalable shape coding process is used for decoding of binary shape.

If spatial scalability is enabled, use_ref_shape is set to 1 and enhancement_type is set to 1, the same non-scalable decoding process is applied for binary shape of enhancement layer. In this case, the following rules are applied for enhancement layer.

1. In PVOP, Inter shape coding should be done as bab_type of co-located MB in the reference VOP (lower layer) is "Opaque".
2. In BVOP, forward reference VOP, most recently decoded non-empty VOP in the same layer, is always selected as reference VOP of shape coding.

If spatial scalability is enabled, use_ref_shape is set to 1 and enhancement_type is set to 0, then the up-sampled binary shape from base layer is used for the binary shape of enhancement layer. The up sampling and down sampling process of this purpose also follows up-down sampling method described in the subclause 1.1.1.4.

When spatial scalability is enabled and enhancement_type is set to 0, in the enhancement layer, the forward prediction in P-VOP and the backward prediction in B-VOP are used as the spatial prediction. In that case the shape information is coded by scan interleaving (SI) based method. For the forward prediction in B-VOP a motion compensated temporal prediction is made from the reference VOP in the enhancement layer. In that case the shape information is coded by the CAE method as like in base layer except that the shape motion vectors are obtained from those of the collocated bab in the lower layer. Motion vector and shape coding mode(bab_type) of collocated bab in the lower layer are used for decoding the enhancement layer bab. The location of collocated bab in the lower layer can be found by following method.

```
    collocated_MBX
        = min ( max ( 0, current_MBX*shape_hor_sampling_factor_m/shape_hor_sampling_factor_n ),
                NumMBXLower-1 );
    collocated_MBY
        = min ( max( 0, current_MBY*shape_ver_sampling_factor_m/shap_ever_sampling_factor_n ),
                NumMBYLower-1);
```

For the current MB location [current_MBX, current_MBY], the location of collocated bab in the base layer is denoted as [collocated_MBX, collocated_MBY]. current_MBX, current_MBY, collocated_MBX and collocated_MBY are the MB-unit coordinations. NumMBXLower and NumMBYLower denote the number of micro-blocks in the lower layer VOP on horizontal and vertical directions, respectively.

1.1.1.3 Prediction in the enhancement layer

A motion compensated temporal prediction is made from the reference VOP in the enhancement layer. In addition, a spatial prediction is formed from the lower reference layer decoded VOP. These predictions are selected individually or combined to form the actual prediction.

1.1.1.4 Spatial prediction

The spatial prediction is made by resampling the lower reference layer reconstructed VOP to the same sampling grid as the enhancement layer. For the resampling, repetition is used on the the lower layer.
For enhancement layer encoding/decoding, the base layer VOP should be up-sampled as the sampling ratio, which is included in the VOL syntax. In VOL syntax for enhancement layer, there are four fields for the up-sampling ratio, i.e., shape_hor_sampling_factor_n, shape_hor_sampling_factor_m, shape_vert_sampling_factor_n and shape_vert_sampling_factor_n.

Down sampling

For the sampling rate, below sampling process is adopted for both horizontal and vertical directions.
For the down-sampling rate m/n,

$m/n = [1/2^k][(m2^k)/n]$, where k is the largest integer that is equal or smaller than $\log_2(n/m)$.

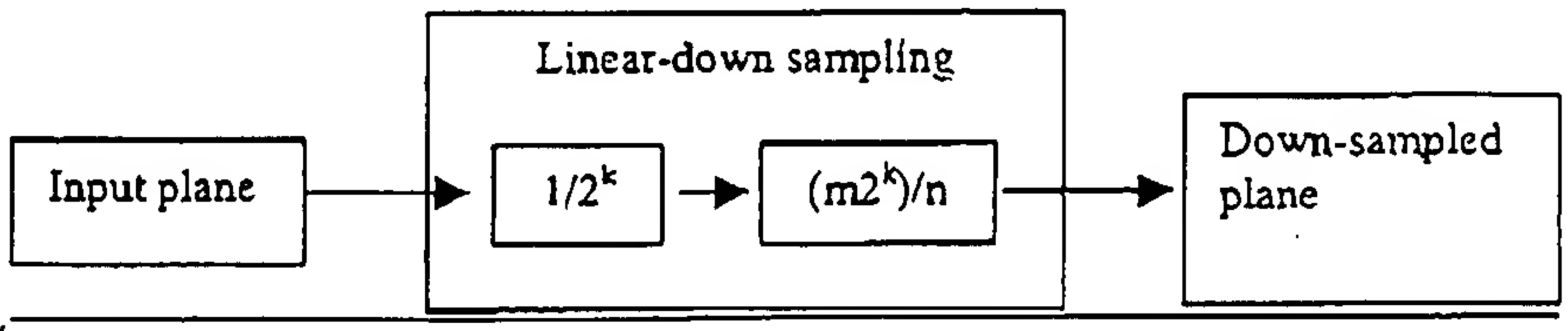


Figure V2 - 1 -- Down sampling process

1. Find $k = \text{floor}(\log_2(n/m))$
2. Sampling a pixel per 2^k pixels :
   if($x\%2^k == (2^k - 1)$) Sampling F(x) pixel ;
3. On the result of 2, sampling a pixel per $n/(m2^k)$ pixels.

Up sampling

For the sampling rate, below sampling process is adopted for both horizontal and vertical directions.
For the up-sampling rate n/m,

$n/m = [n/(m2^k)]\,[2^k]$. where k is the largest integer that is equal or smaller than $\log_2(n/m)$.

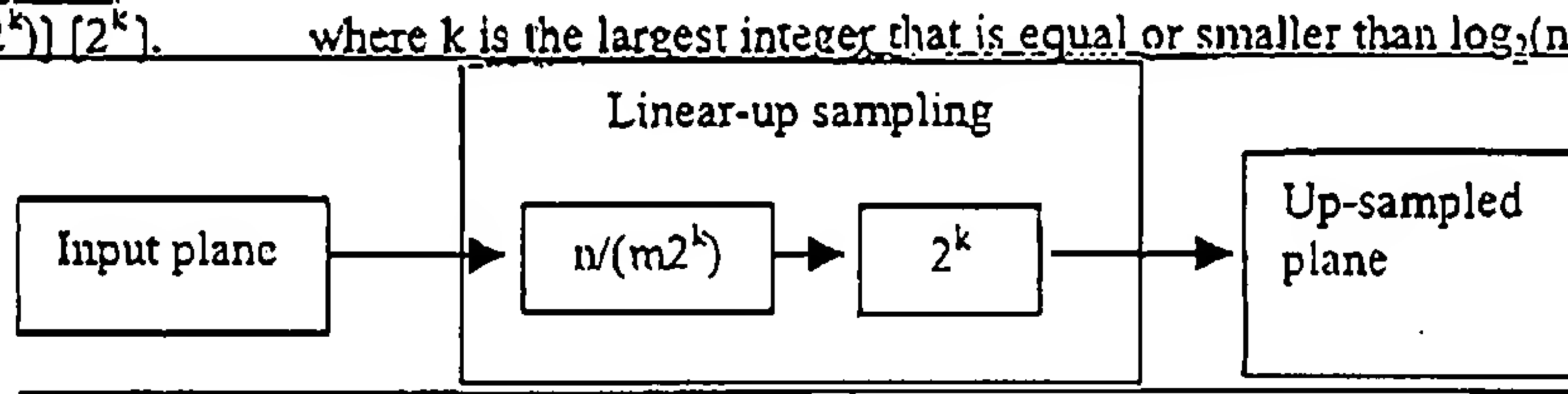


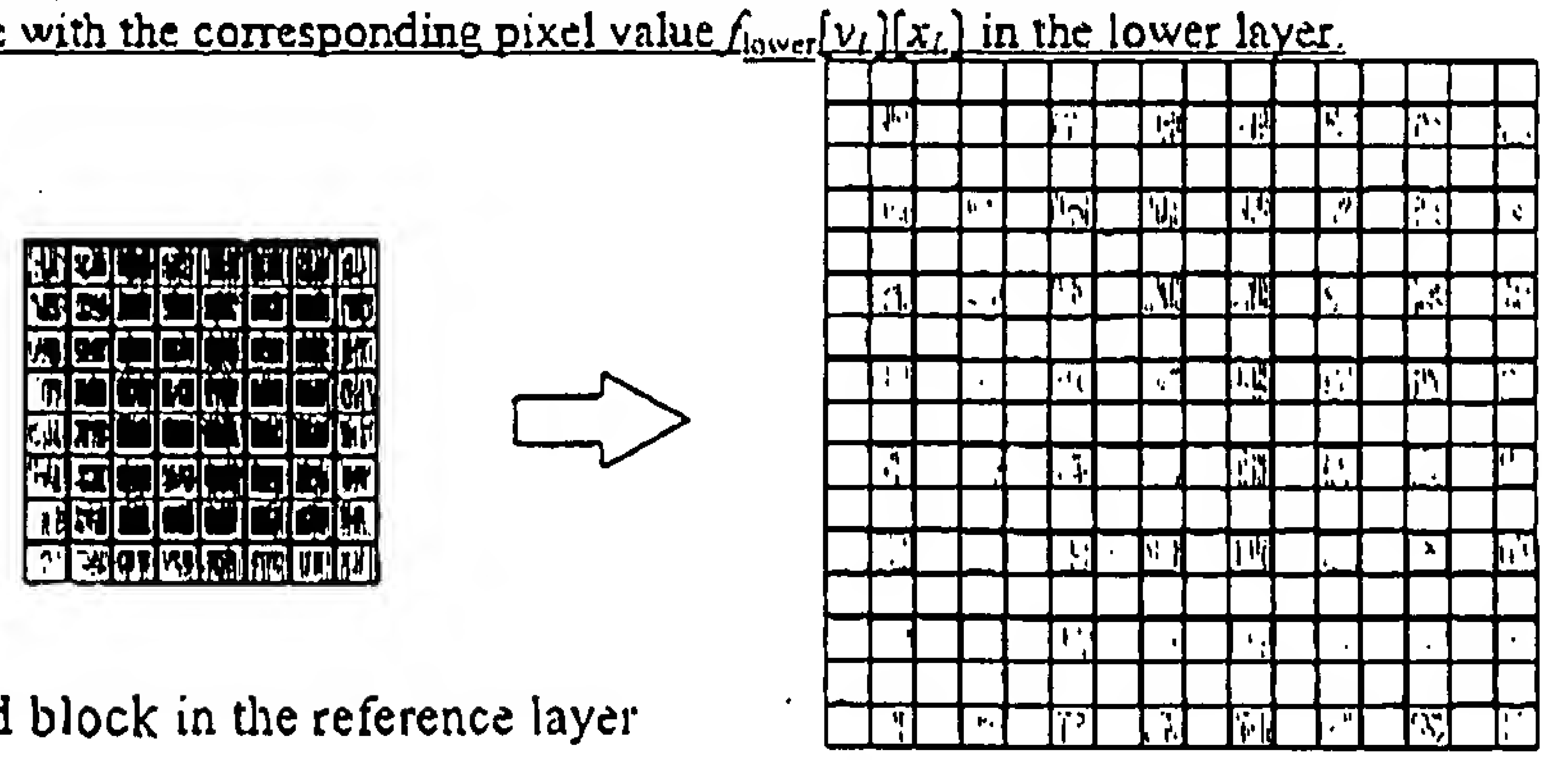
Figure V2 - 2 -- Up sampling process

1. Find $k = \text{floor}(\log_2(n/m))$
2. Repeat one pixel per $(m2^k)/(n-m2^k)$ as follows :
3. On the result of 2, Repeat $2^k - 1$ pixels per pixel.

For a case that the sampling ratio is 1/2 for both horizontal and vertical direction, one pixel value in the lower layer is upsampled to 2x2 pixels in the current layer by the following methods:
The binary alpha image $f_{current}[y_H][x_H]$ in the enhancement layer is upsampled by repetition of the pixels of the image $f_{lower}[y_L][x_L]$ in the lower layer according to the following formula:

$$f_{current}[2y_L+i][2x_L+j] = f_{lower}[y_L][x_L]. \qquad (i, j = 0, 1)$$

In Figure 7-a, the spatial prediction of one BAB is shown for example. In the enhancement layer the pixel value $f_{current}[2y_L+1][2x_L+1]$ is always same with the corresponding pixel value $f_{lower}[y_L][x_L]$ in the lower layer.



Collocated block in the reference layer

Upsampled BAB for spatial prediction

Figure V2 - 3 -- Asso[illegible]ed pixel locations and spatial prediction [illegible]urrent BAB by repetition

In the case of arbitrary shape VOPs, its size and locations change from time to time. In order to ensure the formation of the spatial prediction, it is necessary to identify the location of the resampled VOPs in the absolute coordinate system. VOP_horizontal_mc_spatial_ref and VOP_vertical_mc_spatial_ref in the resampled reference layer should be coincident with those in the enhancment layer. Hence after the position of the spatial reference of the reference layer is scaled by * sampling_factor_n/* sampling_factor_m, the resampled reference VOP should be relocated by referencing the offset values that are the differences between the positions of the spatial references of two VOPs in the horizontal and the vertical direction. Since the shape motion vectors and the shape modes in the reference layer are also needed for the actual shape coding in the enhancement layer bab, the bab type matrix and the motion vector matrix are resampled by repetition. In case of the motion vector matrix each components are scaled by the sampling ratio on both directions.

1.1.1.5 Temporal prediction

A motion compensated temporal prediction is made from the reference VOP in the enhancement layer. For predicting a bab, the motion vectors of the collocated bab in the lower layer is used. Enhancement layer shape motion vector [mvX, mvY] can be used by scaling the lower layer shape motion vector [mvXLower, mvYLower]. The scaling method is as follows:

mvX = (mvXLower * shape_hor_sampling_factor_n) // shape_hor_sampling_factor_m
mvY = (mvYLower * shape_ver_sampling_factor_n) // shape_ver_sampling_factor_m

1.1.1.6 Types for enhancement layer bab

Each bab belongs to one of four types given in Table V2 - 1. This enh_bab_type field influences decoding of further shape information of enhancement layers. For P-VOP coding only 0 and 1 of the enh_bab_type are allowed. For B-VOP coding scheme, all of the four modes are allowed.

Table V2 - 1 -- List of enh_bab_types and usage

enh_bab_type	Semantic	Used in
0	intra NOT_CODED	P-/B- VOPs
1	intra CODED	P-/B- VOPs
2	inter NOT_CODED	B- VOPs
3	inter CODED	B- VOPs

When the enh_bab_type == 1, the bab is predicted by upsampling from the collocated block in the lower reference layer and is decoded by intra-mode CAE based on Scan Interleaving(SI).

The enh_bab_type == 0 means that the error of the predicted bab from the lower reference layer is less than the threshold. The predicted bab is given by upsampling the collocated block in the lower reference layer. The spatial prediction method in 7.5.4.4 or linear repetition method can be used for the upsampling. The predicted bab is used as the current bab.

For enh_bab_type==2 and enh_bab_type==3, no motion vector is decoded for the temporal prediction of enhancement layer bab. The motion vector of the collocated block in the lower reference layer can be utilized. Each component of the vector are scaled by two for referencing. If the collocated bab in the lower reference layer is intra coding mode, the value of the motion vector is set to zero. If enh_bab_type is decode as "inter NOT_CODED", then the error of the predicted bab from the previous VOP is less than the threshold. And the predicted bab is used as the current bab. If enh_bab_type is decode as "inter CODED", the current bab is decoded by inter-CAE which is used in non-scalable shape decoding.

1.1.1.7 Decoding bab types for enhancement layer

In order to decode the bab types, the collocated bab of the lower layer is also examined. Based on the lower layer bab types, the following VLC table is used for decoding the enh_bab_type field. For B-VOP decoding in enhancement layer enh_bab_type is decoded using in Table V2 - 3. For P-VOP decoding in enhancement layer, there are only two enh_bab_type fields, i.e., 0 and 1, and they are decoded by 1~2 bit described in Table V2 - 2.

Table V2 - 2 -- Decoding of enh_bab_type for P-VOP in enhancement layer

enh_bab_type	Code
0	1
1	01

Table V2 - 3 -- Decoding of enh_bab_type for B-VOP in enhancement layer

		enh_bab_type in the enhancement layer			
		0	1	2	3

enh_bab_type [illegible]e lower reference [illegible]er or (bab_type in the base layer)	0(2,3)	1	01	001	000
	1(4)	110	0	10	111
	2(0,1)	001	01	1	000
	3(5,6)	110	0	111	10

Note that the numbers 0-3 in these tables correspond to the enh_bab_types given in 7.8.2.1.6. When the lower layer is the base layer, {"MVDs==0 && No Update", "MVDs!=0 && No Update"}, {"all_0", "all_255"}, {"intraCAE"} and {"MVDs==0 && interCAE", "MVDs!=0 && interCAE"} modes are converted to the enh_bab_type 2, 0, 1 and 3, respectively.

1.1.1.8 Intra coded enhancement layer decoding

Intra coded enhancement layer decoding uses scan interleaving algorithm before performing intra-mode CAE. The decoding order with SI scanning is as follows :

1. Copy B from Base layer
2. Decoding order with Vertical scanning : Vr --> Vp1 --> Vp2 --> ... --> Vpk
3. Decoding order with Horizontal scanning : Hr --> Hp1 --> Hp2 --> ... --> Hpl

where,

B : Piexl that can be copied from collocated pixel in the base layer.

V_r: Vertical scanning Pixel of Residual term of vertical linear-up sampling

V_{pk}: Vertical scanning Pixel of 2^k term of vertical linear-up sampling

H_r: Horizontal scanning Pixel of Residual term of Horizontal linear-up sampling

H_{pl}: Horizontal scanning Pixel of 2^l term of Horizontal linear-up sampling

k : The largest integer that is equal or smaller than

$\log_2$(shape_hor_sampling_factor_n / shape_hor_sampling_factor_m)

l : The largest integer that is equal or smaller than

$\log_2$(shape_ver_sampling_factor_n / shape_ver_sampling_factor_m)

For an example, when shape_hor_sampling_factor_n = 8, shape_hor_sampling_factor_m = 3, shape_ver_sampling_factor_n = 5, shape_hor_sampling_factor_m =1, the scan order of pixels for decoding is shown in Figure V2 - 4.

Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr	Hr
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1	Hp1
Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2	Hp2
Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]	Vp1	[illegible]

Figure V2 - 4 -- An example of decoding order when enh_bab_type==1 (Decoding order: B → V_r → V_{p1} → H_r → H_{p1} → H_{p2}).

1.1.1.9 Border formation fo[illegible]ra-mode CAE for enhancement layer dec[illegible]g

The border formation for decoding of intra CODED bab (enh_bab_type==1) is the same as that of the non-scalable shape coding (see the attached file,Border-formation.doc) except E and U regions in 오류! 참조 원본을 찾을 수 없습니다.. The border of E and U region of the current bordered bab contains the pixels from the collocated regions in the up-sampled lower layer plane and which consists of 2 lines in bottom and right of the current bab.

1.1.1.10 Scan Interleaving Algorithm

The spatial (or resolution) scalability can be easily sought by incorporating the so-called Scan Interleaving (SI) method. In SI, coded-scan-lines are decoded by using the prediction from the closest upper and lower neighboring two lines (reference-scan-lines) as shown in Figure V2 - 5. In the figure, horizontal direction scanning is described. The pixels on coded-scan-lines are the pixels to be decoded and the pixels on reference-scan-lines are the pixels that are already reconstructed The closest upper and lower neighboring pixel values in reference-scan-lines are used for reconstructing the current pixel. (The closest upper and lower neighboring pixels mean the closest left and right neighboring pixels in vertical direction scanning case.) By using those two reference pixels, the current decoded pixel has three types as follows:

Transitional sample data (TSD): The closest upper and lower neighboring pixels that are already reconstructed have different colors each other.

Exceptional sample data (ESD): The closest upper and lower neighboring pixels that are already reconstructed are same and the current pixel is different with the neighboring pixels.

Predictable sample data (PSD): The closest upper and lower neighboring pixels that are already reconstructed are the same as the current pixel.

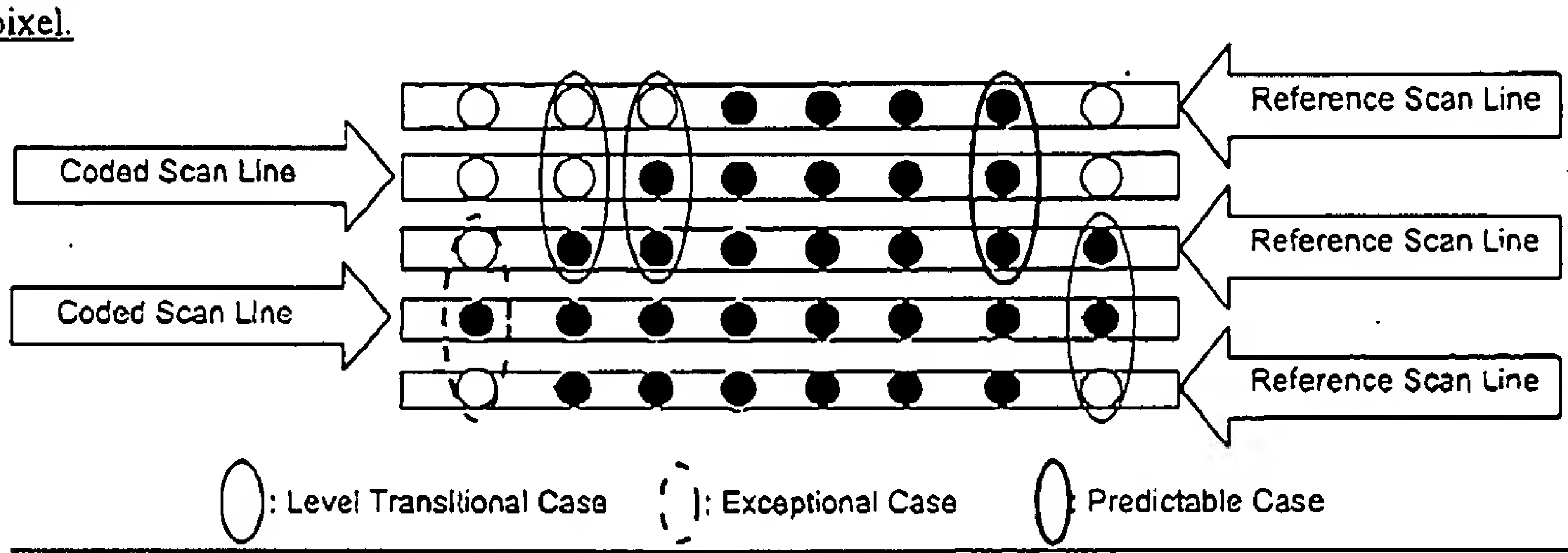


Figure V2 - 5 -- Scan Interleaving for Scalable Coding: an example of a times by 2 case.

To encoding/decoding shape mask plane, SI performs the scanning method in vertical and horizontal directions, successively. SI with vertical scanning (V-SI) and SI with horizontal scanning (H-SI) are separately applied to the input image.

For decoding a bab in the current layer, the spatially predicted bab from collocated block in the lower layer is enhanced by V-SI. And then H-SI is successively performed on the enhanced plane of the V-SI. The reconstructing order is presented in subclause 1.1.1.8.

In the first step of decoding a bab, a binary flag, which denotes the current bab type of SI, is decoded by BAC. There are two bab types of SI : transitional bab (TSD-B) and exceptional bab (ESD-B). If a bab has no ESD, the bab is set to TSD-B. Otherwise the bab is set to ESD-B. When the flag is '0', the current bab is TSD-B. Otherwise, the current bab is ESD-B.

In the encoder, for encoding TSD-B, only TSD pixel are encoded by CAE with the context shown in 오류! 참조 원본을 찾을 수 없습니다.. And for encoding ESD-B, all the pixels in the coded-scan-lines are encoded by CAE with the context shown in 오류! 참조 원본을 찾을 수 없습니다.. The encoding order is the same as shown in subclause 1.1.1.8.

If a bab is TSD-B, only TSD are decoded in the order of V-SI and H-SI. The specific decoding order is presented in 1.1.1.8. The decoding process of TSD-B is as follows.

Following the scanning order, if the current pixel is TSD, then the pixel is decoded by BAC with the context in 오류! 참조 원본을 찾을 수 없습니다.. And if the two neighbor pixels (the closest upper(left) and bottom(right)) are the same value, then the current pixel has the same value as its neighbor pixels.

If a bab is ESD-B, all the pixels in coded scan line are decoded by BAC with the context in 오류! 참조 원본을 찾을 수 없습니다..

1.1.1.11 Determination of the scanning type in enhancement layer decoding.

When enh_bab_type is 1 (intra CODED) and the scale ratio between the layers is square(i.e 200x300 <--> 400x600), before performing BAC, the scanning type should be determined. For the determination, the top and left borders of the collocated block of the lower layer are extended with its neighboring pixels like Figure V2 - 6. If an extended border is outside VOP, the pixel values of the extended border are set to zero. Two counts, N_L and N_U are calculated (N_L: the number of pixels which have different values from the left neighboring pixel value; N_U: the number of pixels which have different values from the upper neighboring value). If N_U is greater than N_L, the spatially predicted bab for decoding the enhancement layer bab has to be

transposed and then the intra-mode CAE for enhancement layer is performed. Otherwise, the intra-mode CAE for enhancement layer is performed on the spatially predicted bab without transposing, that is default condition.

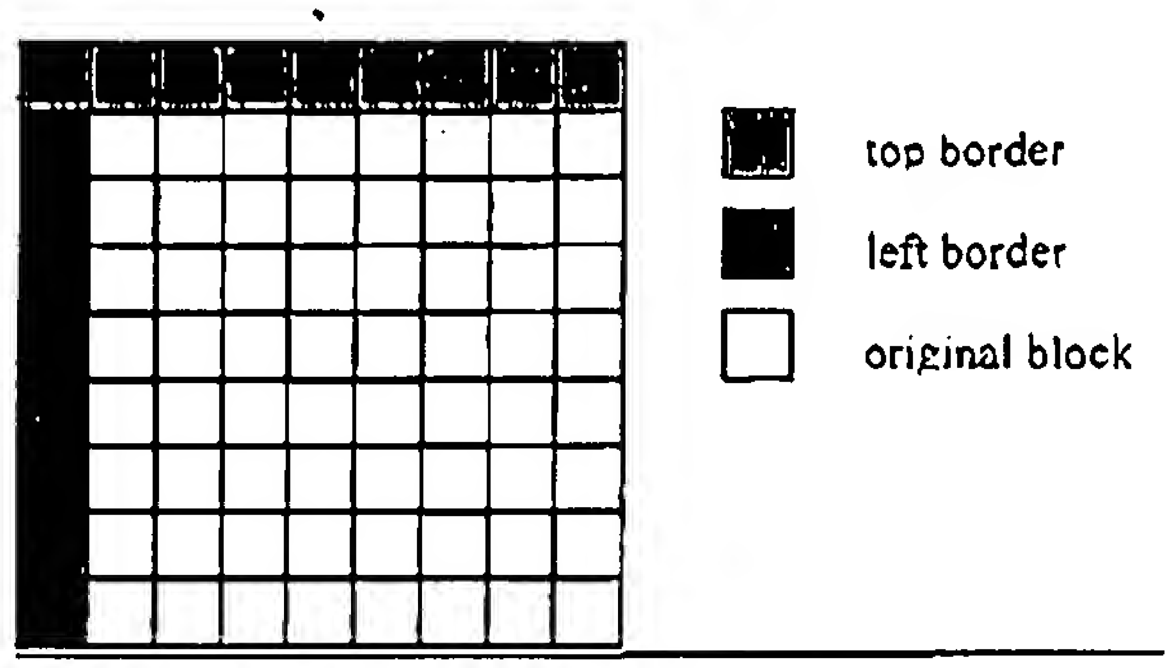


Figure V2 - 6 -- Extension of border pixels.

1.1.1.12 Decoding of bab type of SI

The first decoded symbol of enh_binary_arithmetic_code(), which performs BAC, denotes the current bab type of SI. The probability of SI_bab_type_prob[] given in Annex B is used for decoding bab type of SI. When this flag is '0', the current bab is TSD-B. Otherwise, the current bab is ESD-B.

1.1.1.13 Decoding of intra-mode CAE for enhancement layer

For decoding intra mode CAE for enhancement layer, BAC using the V-SI and H-SI decoding method is performed. By using neighboring 7 bits context, a pixel is arithmetic decoded as in CAE, such that:

Compute a context using the templates shown in 오류! 참조 원본을 찾을 수 없습니다..((a) : for V-SI, (b): for H-SI)

Using the context, find probability from the probability table, i.e. enh_intra_v_prob[] for V-SI and enh_intra_h_prob[] for H-SI. These tables are given in Annex B.(see the attached file, Probability-table)

Decoding the pixel value using the probability.

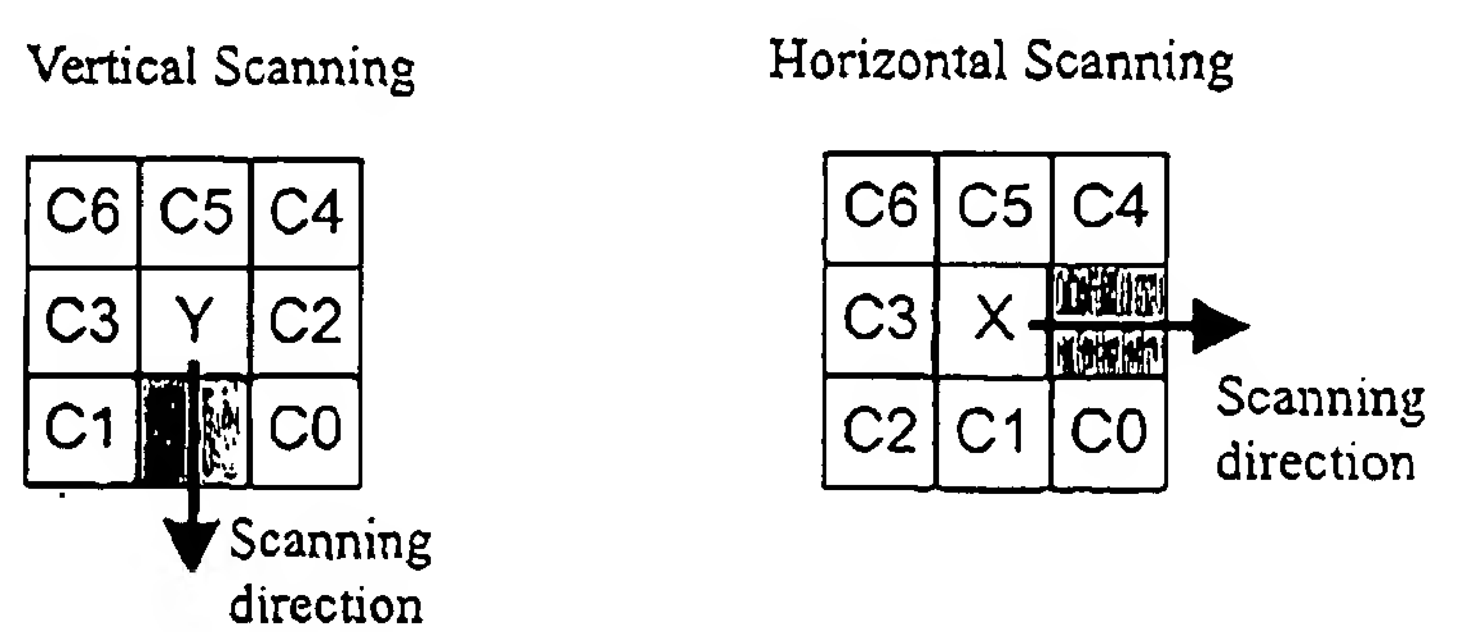
The following subclause describes the computation of the contexts for intra mode CAE.

1.1.1.14 Contexts for intra-mode CAE for enhancement layer

Depending on the scanning direction, a 7 bit context $C = \sum_k c_k \cdot 2^k$ is built for each coded pixel as shown in Figure V2 - 7.

where c_k=0 for transparent pixels and c_k=1 for opaque pixels. (a) of Figure V2 - 7 is for the intra-mode CAE with V-SI (SI by the vertical scanning), and (b) of Figure V2 - 7 is for the intra-mode CAE with H-SI (SI by the horizontal scanning). In (a) of the figure, C6, C5 and C4 are the closest decoded pixels in top-left, top, and top-right location, respectively. C3, and C2 are the closest decoded pixels in left and right location, respectively. And C1 and C0 are the closest decoded pixels in bottom-left and bottom-right location, respectively.

In (b) of the figure, C6, C5 and C4 are the closest decoded pixels in top-left, top, and top-right location, respectively. C3 is the closest decoded pixel in left location. And C2, C1 and C0 are the closest decoded pixels in bottom-left, bottom, and bottom-right location, respectively.



(a) (b)

Figure V2 - 7 -- The context information at different scanning of SI.

There are some special cases to consider:

- When building contexts, any pixels outside the bounding box of the current VOP to the left and above are assumed to be zero (transparent).
- When building contexts and error_resilient_disable==0, any pixels outside the space of the current video packet to the left and above are assumed to be zero (transparent).
- When constructing the context, any pixel outside the bordered BAB is assumed to be the same pixel value of the closest pixel in the border of the current BAB from the context location.

1.1.1.15 Decoding of inter-mode CAE for enhancement layer

For decoding of enhancement layer with B-VOP coding, the inter-mode CAE is performed in the case of the "inter coded" mode (enh_bab_type==3). The context information in Figure V2 - 8 is utilized for this decoding. The decoding process of a pixel is done in the same way as non-scalable binary shape coding.
When constructing the Inter-mode CAE context the following conditional assignment is performed.

if(c1 is unknown) c1=c2

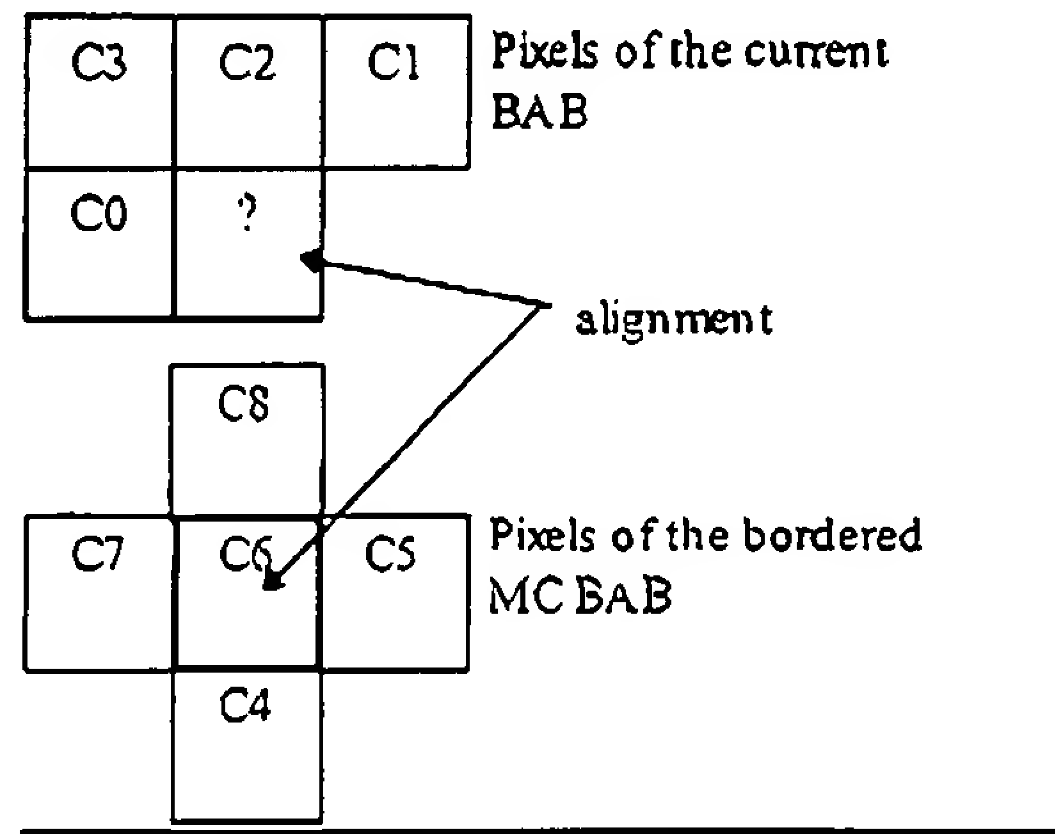


Figure V2 - 8 – The templates of Inter-mode CAE for enhancement layer. C6 is aligned with the pixel to be decoded. The pixel to be decoded is marked with '?'.

1.1.1.16 Quality (SNR) Scalable Coding

Quality Scalability can be easily sought by spatial scalability. Using the spatial scalability and spatial prediction, binary shape can achieve the quality (SNR) scalability. In the decoding process, each spatial scalable layer can be spatially predicted to the full resolution. For the binary shape, this scalable structure is equivalent to quality scalable structure.

1.1.2 Multiple Auxiliary Component Decoding

Auxiliary components are defined for the VOP on a pixel-by-pixel basis, and contain data related to the video object, like disparity, depth, additional texture. Up to 3 auxiliary components (including the grayscale shape) are possible. The number and type of these components is indicated by the video_object_layer_shape_extension given in 오류! 참조 원본을 찾을 수 없습니다.. For example, a value '0000' indicates the grayscale (alpha) shape. The same support region as described in 7.5.4 is used for all auxiliary components, and the decoding procedure is the same as described in subclauses 1.1.1.1 - 1.1.1.6.